



CONVERGENCE RATE OF THE LIMIT THEOREM OF A GALTON-WATSON TREE WITH NEUTRAL MUTATIONS

Xinxin Chen

► To cite this version:

Xinxin Chen. CONVERGENCE RATE OF THE LIMIT THEOREM OF A GALTON-WATSON TREE WITH NEUTRAL MUTATIONS. *Statistics and Probability Letters*, 2013, 83 (2), pp.588-595. hal-00569369

HAL Id: hal-00569369

<https://hal.science/hal-00569369>

Submitted on 24 Feb 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CONVERGENCE RATE OF THE LIMIT THEOREM OF A GALTON-WATSON TREE WITH NEUTRAL MUTATIONS

XINXIN CHEN

Laboratoire des Probabilités et Modèles Aléatoires

Université Pierre-et-Marie-Curie (Paris 6)

4 place Jussieu, 75005 Paris, France

luciechen1988@gmail.com

ABSTRACT. We consider a Galton-Watson branching process with neutral mutations (infinite alleles model), and we decompose the entire population into sub-families of individuals carrying the same allele. Bertoin[3] has established the description of the asymptotic shape of the process of the sizes of the allelic sub-families when the initial population is large and the mutation rate small. The limit in law is a certain continuous state-space branching process (CSBP). Further, we consider the difference between the rescaled size of sub-families and corresponding number of mutations. We find out that it converges in law to some normal distribution with mean zero and whose variance is associated with the CSBP.

Keywords : Branching process, neutral mutations, Lévy-Itô decomposition, weak convergence, Donsker's invariance principle.

1. INTRODUCTION

We consider a Galton-Watson process([1]), that is, a population model with asexual reproduction such that at every generation, each individual gives birth to a random number of children according to a fixed offspring distribution and independently of the other individuals in the population. In this paper, we are interested in the situation where a child can be either a clone, that is, of the same genetic type as its parent, or a mutant, that is, of a new genetic type different from its parent. We stress that each mutant has a distinct allele and in turn gives birth to clones of itself and to new mutants according to the same statistical law as its parent, even though it bears a different allele. In other words, we are working with an infinite alleles model where mutations are neutral for the population dynamics.

To simplify the model, we decompose the entire population into clusters(:sub-families) of individuals having the same allele. This partition will be referred to as the *allelic partition*. In the study of random population models with mutations there are many questions concerning statistics of this allelic partition: what is the probability of observing allelic clusters of certain sizes? how to describe the structure of the allelic partition of the entire population? etc. The paper[2] contains

Date: February 11, 2011.

Key words and phrases. Branching process, neutral mutations, Lévy-Itô decomposition, weak convergence, Donsker's invariance principle.

some answers to these questions. However, our main purpose here is to investigate asymptotical behaviors in law when the size of the population is large (typically as the number of ancestors is large) and mutations are rare. As shown in [3], under some conditions, a non-degenerate limit exists and is conveniently described in terms of a certain continuous state-space branching process in discrete time (CSBP[5]).

Let us show a rough idea here. We consider a fixed reproduction law which is critical and has finite variance, and assume that the Galton-Watson process starts from n ancestors with the same genetic type. We also suppose that neutral mutations affect each child with probability $1/n$. Recall that such a Galton-Watson process becomes extinct after roughly n generations, and that the total population is of order n^2 . So there are only a few mutations at each generation and thus about n different alleles; furthermore the largest allelic sub-family is of order n^2 and the allelic type of mutants from this sub-family (outer degree) is of order n . It is natural to consider the asymptotic features of the rescaled size of the allelic partition.

We use the universal tree \mathbb{U} , which is the set of finite sequences of integers (with \emptyset as the root) to record the genealogy of alleles, and define the tree of alleles as a random process (\mathcal{A}, d) on \mathbb{U} , such that each allele represents a vertex of \mathbb{U} and that the values at vertices are given by the sizes of the corresponding allelic sub-families and the outer degrees, with the convention that the sizes are ranked in the decreasing order for each sibling.

When the size of ancestors is of order n and the rate of mutations is of order $1/n$, we denote by $(\mathcal{A}^{(n)}, d^{(n)})$ the corresponding tree of alleles. Then as n goes to infinity, $n^{-2}\mathcal{A}^{(n)}$ and $n^{-1}d^{(n)}$ converge in the sense of finite dimensional distributions towards the same limit (removing a constant factor). The limit describes the genealogy of a CSBP in discrete time, whose law only depends on the variance of the offspring distribution of the Galton-Watson process. Then we investigate the difference between $n^{-2}\mathcal{A}^{(n)}$ and $n^{-1}d^{(n)}$ multiplied \sqrt{n} , and we show that it converges in law to a "normal" distribution with mean zero. The variance is given by the CSBP.

The plan of this paper is as follows. In section 2, we present precisely the model and our limit theorems. In section 3, we construct the probability structure of the tree of alleles from the random walk. Based on this construction, we describe the ideas to figure out the limit theorems in the paper[3]. In section 4, we prove one central limit improvement of the limit theorems.

2. THE MODEL OF THE TREE OF ALLELES

Our basic data is provided by a pair of non-negative integer-valued random variables

$$\xi = (\xi^{(c)}, \xi^{(m)})$$

which describes the number of clone-children and the number of mutant-children of a typical individual. We are interested in a special situation which appears commonly as a model in population genetics, namely where mutations affect each child according to a fixed probability and independently of the other children (in other words, the conditional distribution of $\xi^{(m)}$ given $\xi^{(c)} + \xi^{(m)} = \ell$

is binomial with parameter (ℓ, p) . We define:

$$\xi^{(+)} = \xi^{(c)} + \xi^{(m)}$$

whose law is noted by $\pi^{(+)}$. We assume that

$$\mathbb{E}[\xi^{(+)}] = 1 \text{ and } \text{Var}(\xi^{(+)}) = \sigma^2 < \infty.$$

We further implicitly exclude the degenerate cases when $\xi^{(c)} = 0$, or $\xi^{(m)} = 0$. For every integer $a \geq 1$, we denote by \mathbb{P}_a the law of a Galton-Watson process with neutral mutations, starting from a ancestors carrying the same genetic type and with reproduction law given by that of $\xi = (\xi^{(c)}, \xi^{(m)})$.

Moreover, we use the notation \mathbb{P}_a^p for the probability measure under which the Galton-Watson process has a ancestors and the mutation rate is p . $\mathcal{L}(\cdot, \mathbb{P}_a^p)$ will then refer to the distribution of a random variable or a process under \mathbb{P}_a^p .

We now take into account mutations by assigning marks to the edges between parents and their mutant children. Since we are interested in the genealogy of alleles, it is convenient to say that an individual has the k -type if its genotype has been affected by k mutations, that is if its ancestral line comprises exactly k marks. We denote by T_k the total population of individuals of the k -th type and by M_k the total number of mutants of k -th type, with the convention that mutants of the 0-th type are the ancestors, i.e. $\mathbb{P}_a(M_0 = a) = 1$.

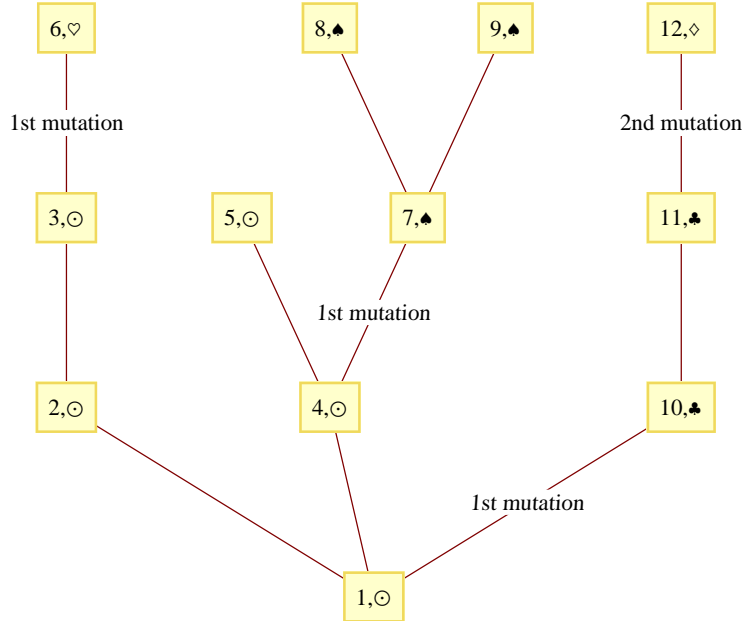


FIGURE 1. The Galton-Watson process with neutral mutations. The symbols $\odot, \heartsuit, \clubsuit, \diamond, \spadesuit$ represent the different alleles. At the same time we enumerate all the vertices.

In order to describe the genealogy of allelic sub-families as random processes indexed by the universal tree, we introduce the set of finite sequences of positive integers

$$\mathbb{U} := \bigcup_{k \in \mathbb{Z}_+} \mathbb{N}^k.$$

where $\mathbb{N} = \{1, 2, \dots\}$ and $\mathbb{N}^0 = \{\emptyset\}$. Let us recall some standard notations in this setting: if $u = (u_1, \dots, u_k)$ is a vertex at level $k \geq 0$ in \mathbb{U} , then the children of u are $uj = (u_1, \dots, u_k, j)$ for $j \in \mathbb{N}$. We also denote by $|u|$ the level of the vertex u , with the convention that the root has level 0, i.e. $|\emptyset| = 0$. We now take advantage of the natural tree structure of \mathbb{U} to record the genealogy of allelic sub-families together with their sizes.

In fact, we can construct a process $\mathcal{A} = (\mathcal{A}_u; u \in \mathbb{U})$ from the given Galton-Watson process with neutral mutations.

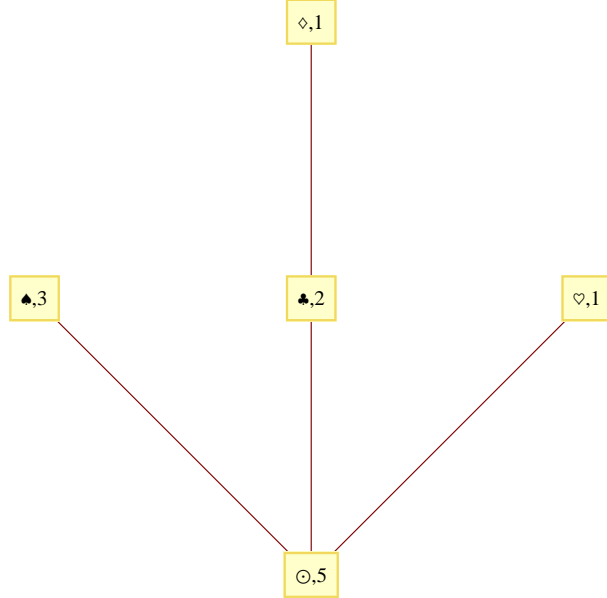


FIGURE 2. The tree of alleles corresponding to the process in Figure 1. The number of each vertex represents the total size of the cluster of its allele.

First, $\mathcal{A}_\emptyset = T_0$ is the size of the sub-family without mutation. Next, recall that M_1 denotes the number of mutants of the first type. We enumerate the M_1 allelic sub-populations of the first type in the decreasing order of their sizes, with the convention that in the case of ties, sub-populations of the same size are ranked uniformly at random. We denote by \mathcal{A}_j the size of j -th allelic sub-population of the first type, agreeing that $\mathcal{A}_j = 0$ if $j > M_1$. We then complete the construction at the next levels by iteration in an obvious way. Specially, if $\mathcal{A}_u = 0$ for some $u \in \mathbb{U}$, then $\mathcal{A}_{uj} = 0$ for all $j \in \mathbb{N}$. Otherwise, we enumerate in the decreasing order of their sizes the allelic sub-populations of type $|u| + 1$ which descend from the allelic sub-family indexed by the vertex u ,

and then \mathcal{A}_{uj} is the size of this j -th sub-family. We call the process $\mathcal{A} = (\mathcal{A}_u; u \in \mathbb{U})$. We define the outer degree of the tree of alleles \mathcal{A} at some vertex $u \in \mathbb{U}$ as

$$d_u := \max\{j \geq 1 : \mathcal{A}_{uj} > 0\}.$$

where we agree that $\max \emptyset = 0$. In words, d_u is the number of allelic sub-populations of type $|u| + 1$ which descend from the allelic sub-family indexed by the vertex u ; in particular, $d_\emptyset = M_1$.

We observe that

$$T_k = \sum_{|u|=k} \mathcal{A}_u \text{ and } M_{k+1} = \sum_{|u|=k} d_u.$$

We construct a tree-indexed random process in the following definition.

Definition 1. Fix $x > 0$ and ν a measure on $(0, \infty)$ with $\int (1 \wedge y) \nu(dy) < \infty$. A tree-indexed CSBP with reproduction measure ν and initial population of size x is a process $(\mathcal{Z}_u; u \in \mathbb{U})$ with values in \mathbb{R}_+ and indexed by the universal tree, whose distribution is characterized by induction on the levels as follows:

- (1) $\mathcal{Z}_\emptyset = x$ a.s.;
- (2) for every $k \in \mathbb{Z}_+$, conditionally on $(\mathcal{Z}_v; v \in \mathbb{U}, |v| \leq k)$, the sequences $(\mathcal{Z}_{uj})_{j \in \mathbb{N}}$ for the vertices $u \in \mathbb{U}$ at generation $|u| = k$ are independent, and each sequence $(\mathcal{Z}_{uj})_{j \in \mathbb{N}}$ is distributed as the family of the atoms of a Poisson random measure on $(0, \infty)$ with intensity $\mathcal{Z}_u \nu$, where atoms are repeated according to their multiplicity, ranked in the decreasing order, and completed by an infinite sequence of 0 if the Poisson measure is finite.

Further, it follows from the definition that $(\sum_{|u|=k} \mathcal{Z}_u; k \in \mathbb{Z}_+)$ is a CSBP in discrete time, with reproduction measure ν and initial population of size x .

In fact, if we define $Z_k := \sum_{|u|=k} \mathcal{Z}_u$, for every $k \in \mathbb{Z}_+$, then

$$Z_0 = \mathcal{Z}_\emptyset = x, \quad \text{a.s.}$$

and for $q \geq 0$

$$\begin{aligned} \mathbb{E}[e^{-qZ_1} | Z_0 = x] &= \mathbb{E}[\exp(-q \sum_j \mathcal{Z}_j)] \\ &= \exp\left(-x \int_{(0, \infty)} (1 - e^{-qy}) \nu(dy)\right) \\ &= \exp(-x\kappa(q)), \end{aligned}$$

where

$$(1) \quad \kappa(q) := \int_{(0, \infty)} (1 - e^{-qy}) \nu(dy).$$

Moreover, the process $(Z_k; k \in \mathbb{Z}_+)$ is a Markov chain with values in \mathbb{R}_+ , starting from x and whose transition probabilities are characterized as follows:

$$\mathbb{E}[e^{-qZ_{k+1}} | Z_k = y] = \exp(-y\kappa(q)).$$

Proposition 1. *If we consider the regime*

$$(2) \quad a(n) \sim nx \text{ and } p(n) \sim cn^{-1} \text{ where } c, n \text{ are some positive constants,}$$

then,

$$(3) \quad \mathcal{L}\left(\{(n^{-2}T_k, n^{-1}M_{k+1}); k \in \mathbb{Z}_+\}, \mathbb{P}_{a(n)}^{p(n)}\right) \Longrightarrow \{(Z_{k+1}, cZ_{k+1}); k \in \mathbb{Z}_+\}$$

where $(Z_k; k \in \mathbb{Z}_+)$ is a CSBP in discrete time with reproduction measure

$$\nu(dy) = \frac{c}{\sqrt{2\pi\sigma^2y^3}} \exp\left(-\frac{c^2y}{2\sigma^2}\right) dy, \quad y > 0,$$

and initial population of size x/c .

Theorem 2. *In the regime 2, the rescaled tree of alleles $(n^{-2}\mathcal{A}_u, u \in \mathbb{U})$ under $\mathbb{P}_{a(n)}^{p(n)}$ converges in the sense of finite dimensional distributions to the tree-indexed CSBP $(\mathcal{Z}_u; u \in \mathbb{U})$ with reproduction measure ν given in the above proposition and random initial population with inverse Gaussian distribution:*

$$\frac{\mathbb{P}(\mathcal{Z}_\emptyset \in dy)}{dy} = \frac{x}{\sqrt{2\pi\sigma^2y^3}} \exp\left(-\frac{(cy-x)^2}{2\sigma^2y}\right) \mathbf{1}_{y>0}.$$

More precisely, if we also take into account the outer degrees $(d_u; u \in \mathbb{U})$, then we have the joint convergence in the sense of finite dimensional distributions holds

$$\mathcal{L}\left(\{(n^{-2}\mathcal{A}_u, n^{-1}d_u); u \in \mathbb{U}\}, \mathbb{P}_{a(n)}^{p(n)}\right) \Longrightarrow ((\mathcal{Z}_u, c\mathcal{Z}_u); u \in \mathbb{U}).$$

The two statements are given and proven by Bertoin in [3].

From Theorem 2, it is immediate that $\frac{c\mathcal{A}_u}{n^2} - \frac{d_u}{n} \Longrightarrow 0$ for any vertex u . Then a natural idea is to study the rate of the convergence.

Theorem 3. *We assume that $p(n) = \frac{c}{n} + o(\frac{1}{n\sqrt{n}})$ and $a(n) \sim nx$, then*

$$\mathcal{L}\left(\{(n^{-2}T_k, \sqrt{n}(\frac{cT_k}{n^2} - \frac{M_{k+1}}{n})); k \in \mathbb{Z}_+\}, \mathbb{P}_{a(n)}^{p(n)}\right) \Longrightarrow \{Z_{k+1}, \mathcal{N}_{cZ_{k+1}}^{(k+1)}; k \in \mathbb{Z}_+\},$$

where $\{\mathcal{N}^{(k)}; k \in \mathbb{N}\}$ is a sequence of independent standard Brownian motions which is independent of $\{Z_k; k \in \mathbb{Z}_+\}$.

Furthermore, assuming that $\{\gamma^{(u)}; u \in \mathbb{U}\}$ is a family of i.i.d. standard BM's which is independent of all random variables mentioned above.

$$\mathcal{L}\left(\left(\frac{A_u}{n^2}; \sqrt{n}\left(\frac{cA_u}{n^2} - \frac{d_u}{n}\right)\right); \mathbb{P}_{a(n)}^{p(n)}\right) \Longrightarrow ((\mathcal{Z}_u, \gamma_{c\mathcal{Z}_u}^{(u)}); u \in \mathbb{U})$$

in the sense of finite dimensional distributions. The law of $\{\mathcal{Z}_u; u \in \mathbb{U}\}$ is described in Theorem 2.

For the sake of complements, we shall prove Proposition 1 and Theorem 2 in Section 3, and the new Theorem 3 in Section 4.

3. THE CONSTRUCTION FROM A RANDOM WALK

We consider a sequence $\{\xi_n = (\xi_n^{(c)}, \xi_n^{(m)}); n \in \mathbb{N}\}$ of i.i.d. variables with law π , and then the random walk starting from $a \geq 1$ and with steps $\xi^{(c)} - 1$,

$$S_k^{(c)} := a + \xi_1^{(c)} + \cdots + \xi_k^{(c)} - k, \quad k \in \mathbb{Z}_+.$$

We still use the notation \mathbb{P}_a for the law of $(S_k^{(c)}; k \in \mathbb{Z}_+)$. We define the first hitting times for the random walk

$$\varsigma(j) := \inf\{k \in \mathbb{Z}_+; S_k^{(c)} = -j\}, \quad j \in \mathbb{Z}_+,$$

and

$$\Sigma(j) := \sum_{i=1}^{\varsigma(j)} \xi_i^{(m)}.$$

Notice that

$$\mathbb{E}[\xi^{(c)}] < 1 \Longrightarrow S_k^{(c)} \rightarrow -\infty \text{ as } k \rightarrow \infty.$$

Hence $\varsigma(j) < \infty$ a.s.

Now we introduce the next lemma which can be thought of as an extension of the well-known result of Otter and Dwass (Section 6.2 in [7]) which associates the distribution of the total population in a (sub-)critical Galton-Watson process with the first hitting time of zero by a random walk. Moreover, this idea will play an important role in the following arguments.

Lemma 1. *Under \mathbb{P}_a ,*

$$\mathcal{L}(T_0, M_1) = \mathcal{L}(\varsigma(0), \Sigma(0)).$$

This conclusion which could be referred to [3] is established by the calculation of the generation function. More precisely, according to the Lagrange inversion formula (Section 5.1[8]), the two dimensional distribution can be given in terms of the offspring distribution.

We denote by $(\mathcal{G}(n); n \in \mathbb{N})$ the natural filtration generated by the sequence $(\xi_n; n \in \mathbb{N})$.

Next, we set $\widetilde{T}_0 := \varsigma(0)$, $\widetilde{M}_1 := \Sigma(0)$ and define for every $k \in \mathbb{N}$ by an implicit recurrence

$$\widetilde{T}_0 + \cdots + \widetilde{T}_k = \varsigma(\widetilde{M}_1 + \cdots + \widetilde{M}_k),$$

and

$$\widetilde{M}_1 + \cdots + \widetilde{M}_{k+1} = \Sigma(\widetilde{M}_1 + \cdots + \widetilde{M}_k) = \sum_{i=1}^{\widetilde{T}_0 + \cdots + \widetilde{T}_k} \xi_i^{(m)}.$$

Corollary 4. *For every $a \geq 1$, the chains $\{(T_k, M_{k+1}); k \in \mathbb{Z}_+\}$ and $\{(\widetilde{T}_k, \widetilde{M}_{k+1}); k \in \mathbb{Z}_+\}$ have the same distribution under \mathbb{P}_a .*

More generally, we can apply the sequence (ξ_n) to construct a random process (\mathcal{A}', d') indexed by the universal tree \mathbb{U} with the same distribution as the tree of alleles (\mathcal{A}, d) . To start with, (\mathcal{A}', d') fulfills the following requirements. First, if $\mathcal{A}'_u = 0$ for some $u \in \mathbb{U}$, then $d'_u = 0$ and $\mathcal{A}'_{uj} = 0$ for all $j \in \mathbb{N}$. Second, for every vertex $u \in \mathbb{U}$ such that $\mathcal{A}'_u > 0$,

$$d'_u = \#\{j \in \mathbb{N} : \mathcal{A}'_{uj} > 0\},$$

which is called the outer degree of \mathcal{A}' at u , is a finite number and $\mathcal{A}'_{uj} > 0$ if and only if $j \geq d'_u$. We set $\mathcal{A}'_\emptyset = \varsigma(0)$ and $d'_\emptyset = \Sigma(0)$. Next, consider the increments

$$\lambda(j) := \varsigma(j) - \varsigma(j-1) \text{ and } \delta(j) := \Sigma(j) - \Sigma(j-1), \quad j \geq 1.$$

For vertices at the first level, $\{(\mathcal{A}'_j, d'_j) : 1 \leq j \leq d'_\emptyset\}$ is given by the rearrangement of the sequence $\{(\lambda(j), \delta(j)); 1 \leq j \leq d'_\emptyset\}$ in the decreasing order with respect to the first coordinate $\lambda(j)$ with the usual convention in case of ties. We may then continue with vertices of the next levels by an iteration which should be obvious.

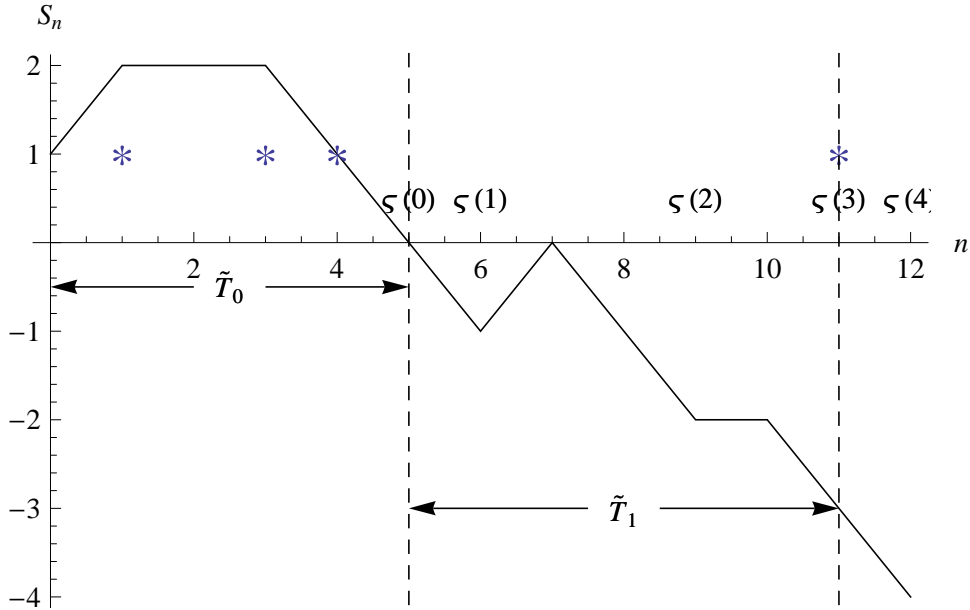


FIGURE 3. The associated random walk with the origin tree shown in Figure 1.

Based on the construction from a random walk, we give the following explanations to prove these limit theorems.

Let $(\xi_k^{(+)}; k \in \mathbb{N})$ be a sequence of i.i.d. copies of $\xi^{(+)}$, then we consider a random walk $(S_k^{(n)}; k \in \mathbb{Z}_+)$ defined by

$$S_0^{(n)} = a(n);$$

and

$$S_k^{(n)} = a(n) + \xi_1^{(+)} + \cdots + \xi_k^{(+)} - k.$$

By Donsker's invariance principle and Skorohod's representation, we may suppose that with probability one

$$\lim_{n \rightarrow \infty} n^{-1} S_{[n^2 t]}^{(n)} = x + \sigma B_t,$$

where $(B_t; t \geq 0)$ is a standard Brownian motion and the convergence holds uniformly on every compact time-interval.

For every fixed n , we now decompose each variable $\xi_i^{(+)}$ as the sum $\xi_i^{(+)} = \xi_i^{(cn)} + \xi_i^{(mn)}$ by using a Bernoulli sampling; that is conditionally on $\xi_i^{(+)} = l$, $\xi_i^{(mn)}$ has the binomial distribution with parameter $(l, p(n))$. We use independent Bernoulli sampling for the different indices i so that the pairs $(\xi_i^{(cn)}, \xi_i^{(mn)})$ are i.i.d. and have the same law as ξ under $\mathbb{P}^{p(n)}$. If we define

$$\begin{aligned} S_k^{(mn)} &:= \xi_1^{(mn)} + \cdots + \xi_k^{(mn)}, \\ S_k^{(cn)} &:= a(n) + \xi_1^{(cn)} + \cdots + \xi_k^{(cn)} - k = S_k^{(n)} - S_k^{(mn)}, \quad k \in \mathbb{Z}_+. \end{aligned}$$

Then

$$\begin{cases} \text{(clone)} & \frac{S_{[n^2 t]}^{(cn)}}{n} \longrightarrow x + \sigma B_t - ct \\ \text{(mutation)} & \frac{S_{[n^2 t]}^{(mn)}}{n} \longrightarrow ct \end{cases}$$

where the convergences hold a.s., uniformly on every compact time-interval.

Now we introduce

$$\varsigma^{(n)}(0) := \inf\{k \in \mathbb{Z}_+; S_k^{(cn)} = 0\} \text{ and } \Sigma^{(n)}(0) := \sum_{i=1}^{\varsigma^{(n)}(0)} \xi_i^{(mn)}.$$

We denote by τ_y the hitting times of a Brownian motion with drift

$$\tau_y := \inf\{t \geq 0; \sigma B_t - ct < -y\}, \quad y \geq 0.$$

Then we get that with probability one

$$\lim_{n \rightarrow \infty} n^{-2} \varsigma^{(n)}(0) = \tau_x \text{ and } \lim_{n \rightarrow \infty} n^{-1} \Sigma^{(n)}(0) = c\tau_x.$$

By application of Lemma 1, we point out that

$$(4) \quad \mathcal{L}\left((n^{-2}T_0, n^{-1}M_1), \mathbb{P}_{a(n)}^{p(n)}\right) \Longrightarrow (\tau_x, c\tau_x).$$

According to the definition of τ , for all $q, t \geq 0$,

$$\mathbb{E}[e^{-q\tau_t}] = \exp\left(-t \frac{\sqrt{c^2 + 2q\sigma^2} - c}{\sigma^2}\right).$$

More precisely, the process $\tau = (\tau_t; t \geq 0)$ is a subordinator with no drift and Lévy measure $c^{-1}\nu$, where ν is given in Proposition 1.

$$\mathbb{E}[e^{-q\tau_x}] = \exp\left(-x \frac{\sqrt{c^2 + 2q\sigma^2} - c}{\sigma^2}\right) = \exp\left(-\frac{x}{c}\kappa(q)\right),$$

where

$$\begin{aligned}\kappa(q) &= \int_{(0,\infty)} (1 - e^{-qy})\nu(dy) \\ &= \int_{(0,\infty)} (1 - e^{-qy}) \frac{c}{\sqrt{2\pi\sigma^2 y^3}} \exp\left(-\frac{c^2 y}{2\sigma^2}\right) dy.\end{aligned}$$

With the subordinator $\tau = (\tau_t; t \geq 0)$ defined as above, we can define a sequence $\zeta = (\zeta_k; k \in \mathbb{Z}_+)$ by implicit iteration as follows:

$$\begin{aligned}\zeta_0 &= \frac{x}{c}, \quad \zeta_1 = \tau_{c\zeta_0}, \quad \zeta_1 + \zeta_2 = \tau_{c(\zeta_0 + \zeta_1)}, \dots, \\ \zeta_1 + \dots + \zeta_{k+1} &= \tau_{c(\zeta_0 + \dots + \zeta_k)} \dots\end{aligned}$$

Observe that for any $q \geq 0$,

$$\mathbb{E}[e^{-q\zeta_1}] = \exp\left(-\frac{x}{c}\kappa(q)\right).$$

By induction, it can be shown that the random times $c(\zeta_0 + \zeta_1 + \dots + \zeta_k)$ are stopping times in the natural filtration of τ . Indeed, first:

$$\{c(\zeta_0 + \zeta_1) \leq t\} = \{x \leq t\} \cap \{\tau_x \leq \frac{t-x}{c}\} \in \sigma(\tau_s; s \leq t) \text{ for every } t \geq 0.$$

Next, assuming that we have proved that $c(\zeta_0 + \zeta_1 + \dots + \zeta_k)$ is a stopping time, then for every $t \geq 0$

$$\begin{aligned}\{c(\zeta_0 + \dots + \zeta_{k+1}) \leq t\} &= \{c(\zeta_0 + \dots + \zeta_k) \leq t\} \cap \{\tau_{c(\zeta_0 + \dots + \zeta_k)} \leq \frac{t-x}{c}\} \\ &\in \sigma(\tau_s; s \leq t).\end{aligned}$$

So we can apply the strong Markov property to obtain:

$$\begin{aligned}\mathbb{E}[e^{-q\zeta_{k+1}} | \zeta_k = y] &= \mathbb{E}[\exp(-q(\tau_{c(\zeta_0 + \dots + \zeta_k)} - \tau_{c(\zeta_0 + \dots + \zeta_{k-1})})) | \zeta_k = y] \\ &= \mathbb{E}[\exp(-q\tau_{cy})] \\ &= \exp(-y\kappa(q)).\end{aligned}$$

Obviously, the sequence ζ is a CSBP with reproduction law ν and initial value x/c . In the same way as above, we can directly show:

$$\mathcal{L}\left((n^{-2}\tilde{T}_k, n^{-1}\tilde{M}_{k+1}) : k \in \mathbb{Z}_+, \mathbb{P}_{a(n)}^{p(n)}\right) \Longrightarrow ((\zeta_{k+1}, c\zeta_{k+1}) : k \in \mathbb{Z}_+).$$

By application of Corollary 4, we get the conclusions of Proposition 1:

$$\mathcal{L}\left((n^{-2}T_k, n^{-1}M_{k+1}) : k \in \mathbb{Z}_+, \mathbb{P}_{a(n)}^{p(n)}\right) \Longrightarrow ((\zeta_{k+1}, c\zeta_{k+1}) : k \in \mathbb{Z}_+).$$

The general branching property by iteration for the original Galton-Watson process, now yields the characterization of the probabilistic structure of the tree of alleles. That is, for every integer $a \geq 1$ and $k \geq 0$, under \mathbb{P}_a , conditionally on $((\mathcal{A}_v, d_v); |v| \leq k)$, the tree of alleles fulfills the following properties:

<1>: the families of variables

$$\{((\mathcal{A}_{uj}, d_{uj}); 1 \leq j \leq d_u) : |u| = k \text{ such that } \mathcal{A}_u > 0\}$$

are independent;

<2>: for each vertex u satisfying that $|u| = k$ and that $\mathcal{A}_u > 0$, the d_u -tuple $((\mathcal{A}_{uj}, d_{uj}); 1 \leq j \leq d_u)$ has the same distribution as $\mathcal{L}((T_0, M_1)^{(d_u)}; \mathbb{P}_1)$.

Given $d_\emptyset = b(n)$ with any positive constant $b > 0$, $((\mathcal{A}_u, d_u); |u| = 1)$ is distributed as the law of the rearrangement of the sequence $\{(\lambda(j), \delta(j)); 1 \leq j \leq b(n)\}$ in the decreasing order with respect to the first coordinate $\lambda(j)$ with the usual convention in case of ties.

While $n^{-2} \sum_{j=1}^{b(n)} \lambda(j) \Longrightarrow \tau_{b+x} - \tau_x =_d \tau_b$. We apply Lévy-Itô decomposition for the subordinator $\{\tau_s; x \leq s \leq x + b\}$. In this way, we obtain a Poisson point process on $(0, \infty)$ with intensity $bc^{-1}\nu$. Let $(\alpha_1(b), \alpha_2(b), \dots)$ stand for the sequence ranked in the decreasing order of the atoms of such a Poisson point process. Let us show the convergence of the Laplace's transforms:

$$\begin{aligned} \mathbb{E}\left[\exp\left(-q \sum_{j=1}^{b(n)} \frac{\lambda(j)}{n^2}\right)\right] &= \left(1 - \int_{(0, \infty)} qe^{-qy} \mathbb{P}[\lambda(j) > n^2 y] dy\right)^{b(n)} \\ &\longrightarrow \mathbb{E}[\exp(-q\tau_b)] = \exp\left(-\frac{b}{c}\kappa(q)\right) \\ &= \exp\left(-\frac{b}{c} \int_{(0, \infty)} qe^{-qy} \bar{\nu}(y) dy\right). \end{aligned}$$

\Longrightarrow

$$\int_{(0, \infty)} e^{-qy} b(n) \mathbb{P}[\lambda(1) > n^2 y] dy \longrightarrow \frac{b}{c} \int_{(0, \infty)} e^{-qy} \bar{\nu}(y) dy \text{ for any } q > 0.$$

with $\bar{\nu}(y) = \nu((y, \infty))$ for any $y > 0$. Observe that $\bar{\nu}$ is a continuous decreasing function. The law of rare events for null arrays (e.g. Theorem 14.18 in [6]) ensures that under $\mathbb{P}^{p(n)}$ and conditionally on $d_\emptyset = b(n) \sim bn$,

$$\left(\frac{\mathcal{A}_u}{n^2}; |u| = 1\right) \Longrightarrow (\alpha_1(b), \alpha_2(b), \dots)$$

in the sense of finite dimensional distributions.

Similarly, the joint convergence in the sense of finite dimensional distributions can be obtained, that is

$$\left(\left(\frac{\mathcal{A}_u}{n^2}, \frac{d_u}{n} \right); |u| = 1 \right) \Longrightarrow \{(\alpha_1(b), c\alpha_1(b)), (\alpha_2(b), c\alpha_2(b)), \dots\}$$

Moreover, the properties $< 1 >$ and $< 2 >$ enable us to establish Theorem 2.

4. THE RATE OF CONVERGENCE

The construction from the random walk is useful to investigate the rate of convergence of $cn^{-2}\mathcal{A} - n^{-1}d \Longrightarrow 0$.

Proof of Theorem 3: Let us continue with the settings of the random walks. Still, with probability one,

$$\frac{S_{\lfloor n^2 t \rfloor}^{(mn)}}{n} \longrightarrow ct \quad \text{and} \quad \frac{S_{\lfloor n^2 t \rfloor}^{(cn)}}{n} \longrightarrow x + \sigma B_t - ct,$$

uniformly on every compact time-interval. For simplification, we first discuss $cn^{-2}T_0 - n^{-1}M_1$, which is distributed as $c\frac{\varsigma^{(n)}(0)}{n^2} - \frac{S_{\varsigma^{(n)}(0)}^{(mn)}}{n}$ with $n^{-2}\varsigma^{(n)}(0) \rightarrow \tau_x$.

We note that $X_t^{(mn)} = \sqrt{n} \left(c\frac{\lfloor n^2 t \rfloor}{n^2} - \frac{S_{\lfloor n^2 t \rfloor}^{(mn)}}{n} \right)$ and $Y_t^{(cn)} = \frac{S_{\lfloor n^2 t \rfloor}^{(cn)}}{n}$ for convenience.

We have the following joint convergence:

$$(5) \quad (X_t^{(mn)}, Y_t^{(cn)})_{t \geq 0} \Longrightarrow \{\beta_{ct}, x + \sigma B_t - ct; t \geq 0\},$$

where β and B are two independent BM(0)'s. The proof of (5) will be presented later.

Recall that $\varsigma^{(n)}(0)$ is a hitting time of the random walk $S^{(cn)}$. It is sufficient to prove that

$$(\sqrt{n}(c\frac{\tilde{T}_0}{n^2} - \frac{\tilde{M}_1}{n}), \frac{\tilde{T}_0}{n^2}) = (X_{\varsigma^{(n)}/n^2}^{(mn)}, \varsigma^{(n)}/n^2) \Longrightarrow (\beta_{c\tau_x}, \tau_x) =_d (\sqrt{c\tau_x}\mathcal{N}_1^{(1)}, \tau_x).$$

Conditionally on $(\tilde{T}_0, \tilde{M}_1)$, precisely given $\tilde{M}_1 = b(n) \sim bn$ with b a positive constant, we consider the random walk with steps $\xi^{(cn)} - 1$ started from 0 and obtain:

$$\tilde{T}_1 = \varsigma^{(n)}(b(n)) - \varsigma^{(n)}(0) \quad \text{and} \quad \tilde{M}_2 = \sum_{k=\varsigma^{(n)}(0)+1}^{\varsigma^{(n)}(b(n))} \xi_k^{(mn)},$$

which are independent of $\mathcal{G}(\varsigma^{(n)}(0))$.

(5) ensures that conditionally on $\tilde{M}_1 = b(n)$,

$$\begin{aligned} (\sqrt{n}(c\frac{\tilde{T}_1}{n^2} - \frac{\tilde{M}_2}{n}), \frac{\tilde{T}_1}{n^2}) &\Longrightarrow (\beta_{c(\tau_{b+x}-\tau_x)}, \tau_{b+x} - \tau_x) \\ &=_{d} (\sqrt{c(\tau_{b+x} - \tau_x)}\mathcal{N}_1^{(2)}, \tau_{b+x} - \tau_x). \end{aligned}$$

In addition, we observe that $\sqrt{n}(c\frac{\tilde{T}_1}{n^2} - \frac{\tilde{M}_2}{n}) = \sum_{k=\varsigma^{(n)}(0)+1}^{\varsigma^{(n)}(b(n))} \left(\frac{c/n - \xi_k^{(mn)}}{\sqrt{n}} \right)$ is independent of $\mathcal{G}(\varsigma^{(n)}(0))$. It follows that $\mathcal{N}_1^{(1)}$ and $\mathcal{N}_1^{(2)}$ are independent normal variables and that they are both independent of the subordinator τ . Finally, by the definition of $(\tilde{T}_k, \tilde{M}_{k+1})$ and Corollary 4, we obtain that

$$\mathcal{L} \left(\left\{ \left(\frac{T_k}{n^2}, \sqrt{n} \left(\frac{cT_k}{n^2} - \frac{M_{k+1}}{n} \right) \right); k \in \mathbb{Z}_+, \mathbb{P}_{a(n)}^{p(n)} \right\} \right) \Longrightarrow \{Z_{k+1}, \mathcal{N}_{cZ_{k+1}}^{(k+1)}; k \in \mathbb{Z}_+\},$$

Notice that

$$\sqrt{n} \left(c \frac{\widetilde{T}_1}{n^2} - \frac{\widetilde{M}_2}{n} \right) = \sum_{j=1}^{d'_0} \left(c \frac{\mathcal{A}'_j}{n^2} - \frac{d'_j}{n} \right), \frac{T_1}{n^2} = \sum_{j=1}^{d'_0} n^{-2} \mathcal{A}'_j;$$

Here given \widetilde{M}_1 (or d'_0), $\{\mathcal{A}'_j, d'_j; 1 \leq j \leq \widetilde{M}_1\}$ is given by the rearrangement of the sequence $\{\lambda(j), \delta(j); 1 \leq j \leq \widetilde{M}_1\}$ in the decreasing order with respect to the first coordinate. Meanwhile, $\{\sqrt{n}(c \frac{\lambda(j)}{n^2} - \frac{\delta(j)}{n}), \frac{\lambda(j)}{n^2}; 1 \leq j \leq \widetilde{M}_1\}$ is actually $\{X_{t_j}^{(mn)} - X_{t_{j-1}}^{(mn)}, t_j - t_{j-1}; t_0 \leq t_1 \leq \dots \leq t_{d'_0}\}$ where $\{t_j = n^{-2}\zeta(j); 1 \leq j \leq d'_0\}$.

Recall that given $n^{-1}d'_0 = b(n) \sim bn$, then $n^{-2}\zeta(d'_0) \rightarrow \tau_{b+x}$ a.s.. Theorem 2 leads to the fact that the rearrangement of the sequence $\{t_j - t_{j-1}; 1 \leq j \leq d'_0\}$ in the decreasing order converges in law to the rearrangement of the family of jump sizes $\{(\tau_y - \tau_{y-}); x \leq y \leq x + b\}$ in the decreasing order which can be viewed as $\{\alpha_1(b) \geq \alpha_2(b) \geq \dots\}$.

Furthermore, the joint convergence (5) tells us that $\{X_{t_j}^{(mn)} - X_{t_{j-1}}^{(mn)}, t_j - t_{j-1}; t_0 \leq t_1 \leq \dots \leq t_{d'_0}\}$ is asymptotically corresponding to $\{\beta_{\tau_y} - \beta_{\tau_{y-}}, \tau_y - \tau_{y-}; x \leq y \leq x + b\}$. The independence between β and B ensures that conditionally on $\{\tau_y - \tau_{y-}; x \leq y \leq x + b\}$, these $\beta_{\tau_y} - \beta_{\tau_{y-}}$ are independent central normal variables with variance $\tau_y - \tau_{y-}$.

Then the rearrangement of the family $\{X_{t_j}^{(mn)} - X_{t_{j-1}}^{(mn)}, t_j - t_{j-1}; t_0 \leq t_1 \leq \dots \leq t_{d'_0}\}$ in the decreasing order of the second coordinate converges in the sense of finite dimensional distributions to $\{\beta_{c\alpha_k(b)}^{(k)}, \alpha_k(b); k \geq 1\}$ where $(\beta^{(k)}, k \geq 1)$ is a sequence of i.i.d BM's which is independent of $\{\alpha_1(b), \alpha_2(b), \dots\}$.

Under the probability $\mathbb{P}_{a(n)}^{p(n)}$, conditionally on $d'_0 = b(n) \sim bn$,

$$\left\{ \sqrt{n} \left(c \frac{\mathcal{A}'_j}{n^2} - \frac{d'_j}{n} \right), \frac{\mathcal{A}'_j}{n^2}; 1 \leq j \leq d'_0 \right\} \Longrightarrow \{\beta_{c\alpha_j(b)}^{(j)}, \alpha_j(b); j \geq 1\}$$

in the sense of finite dimensional distributions.

Then the properties $< 1 >$ and $< 2 >$ entail Theorem 3.

Now it remains to prove (5). We take $0 = s_0 < s_1 < s_2 < \dots < s_N = 1$ and $2N$ real values $\lambda_1, \dots, \lambda_N; \mu_1, \dots, \mu_N$. When n is large enough, we have: $0 = \lfloor n^2 s_0 \rfloor < \lfloor n^2 s_1 \rfloor < \dots < \lfloor n^2 s_N \rfloor = n^2$. Thus, for every integer $j \in \{1, \dots, n^2\}$, there exists $k_j \in \{0, \dots, N\}$ such that $\lfloor n^2 s_{k_j-1} \rfloor < j \leq \lfloor n^2 s_{k_j} \rfloor$. Let $a_j = \lambda_{k_j} + \dots + \lambda_N$ and $b_j = \mu_{k_j} + \dots + \mu_N$. Let us compute the Fourier's transforms:

$$\begin{aligned}
\Lambda_N^{(n)} &:= \mathbb{E} \left[\exp \left(i \sum_{k=1}^N \lambda_k X_{s_k}^{(mn)} + i \sum_{k=1}^N \mu_k Y_{s_k}^{(cn)} \right) \right] \\
&= \mathbb{E} \left[\exp \left(-i \sum_{j=1}^{n^2} \frac{a_j}{\sqrt{n}} \left(\xi_j^{(mn)} - \frac{c}{n} \right) + i \sum_{j=1}^{n^2} \frac{b_j}{n^2} \left(\xi_j^{(cn)} - 1 \right) \right) \right] \\
&= \exp \left(i \sum_{k=1}^N \mu_k \frac{a(n)}{n} \right) \prod_{j=1}^{n^2} g \left[\exp \left(\frac{ib_j}{n} \right), \exp \left(-i \frac{a_j}{\sqrt{n}} \right) \right] \exp \left\{ ia_j \frac{c}{n\sqrt{n}} - ib_j \frac{1}{n} \right\}.
\end{aligned}$$

It follows that

$$\begin{aligned}
\log(\Lambda_N^{(n)}) &\sim \sum_{j=1}^{n^2} \left[\log \left(\frac{ib_j}{n} - \frac{ia_j}{n\sqrt{n}} - \frac{ib_j c + 1/2b_j^2 + 1/2a_j^2 c + 1/2b_j^2 \sigma^2}{n^2} \right) \right. \\
&\quad \left. + \left(i \frac{a_j c}{n\sqrt{n}} - i \frac{b_j}{n} \right) \right] + \left(i \sum_{k=1}^N \mu_k \frac{a(n)}{n} \right)
\end{aligned}$$

Thanks to the independence of the increments of $(X^{(mn)}, Y^{(cn)})$, we only need to check the finite dimensional convergence for one fixed s_1 . Actually,

$$\begin{aligned}
\log(\Lambda_1^{(n)}) &\sim i\mu_1 \frac{a(n)}{n} - \frac{ic}{n^2} \sum_{j=1}^{\lfloor n^2 s_1 \rfloor} \mu_1 + \frac{c}{2n^2} \sum_{j=1}^{\lfloor n^2 s_1 \rfloor} \lambda_1^2 + \frac{\sigma_2}{2n^2} \sum_{j=1}^{\lfloor n^2 s_1 \rfloor} \mu_1^2 \\
&\longrightarrow i\mu_1 x + \frac{\sigma^2 \mu_1^2 s_1}{2} - ic\mu_1 s_1 + \frac{c\lambda_1^2 s_1}{2}
\end{aligned}$$

Hence $\Lambda_1^{(n)} \longrightarrow \exp \left(i\mu_1 x + \frac{\sigma^2 \mu_1^2 s_1}{2} - ic\mu_1 s_1 + \frac{c\lambda_1^2 s_1}{2} \right)$, which is exactly

$$\mathbb{E}[\exp(i\mu_1(x + \sigma B_{s_1} - cs_1))] \mathbb{E}[\exp(i\lambda_1 \beta_{cs_1})].$$

Now we only have to verify the tightness of $(X_t^{(mn)}, Y_t^{(cn)}; 0 \leq t \leq 1)$. We take $0 \leq s \leq r \leq t \leq 1$, then we estimate

$$\Gamma(s, r, t) := \mathbb{E} \left[\left\| (X_t^{(mn)}, Y_t^{(cn)}) - (X_r^{(mn)}, Y_r^{(cn)}) \right\|^2 \right] \mathbb{E} \left[\left\| (X_r^{(mn)}, Y_r^{(cn)}) - (X_s^{(mn)}, Y_s^{(cn)}) \right\|^2 \right].$$

$$\begin{aligned}
& \mathbb{E} \left[\left\| (X_t^{(mn)}, Y_t^{(cn)}) - (X_r^{(mn)}, Y_r^{(cn)}) \right\|^2 \right] \\
&= \mathbb{E} \left[\frac{1}{n} \left(\sum_{k=\lfloor n^2 r \rfloor + 1}^{\lfloor n^2 t \rfloor} (\xi_k^{(mn)} - \frac{c}{n}) \right)^2 \right] + \mathbb{E} \left[\frac{1}{n^2} \left(\sum_{k=\lfloor n^2 r \rfloor + 1}^{\lfloor n^2 t \rfloor} (\xi_k^{(cn)} - 1) \right)^2 \right] \\
&= \frac{\lfloor n^2 t \rfloor - \lfloor n^2 r \rfloor}{n} \left(1/2\sigma^2 p(n)^2 + p(n)(1 - p(n)) \right) + \left(\frac{\lfloor n^2 t \rfloor - \lfloor n^2 r \rfloor}{\sqrt{n}} (p(n) - c/n) \right)^2 \\
&+ \frac{\lfloor n^2 t \rfloor - \lfloor n^2 r \rfloor}{n^2} \left(1/2\sigma^2 (1 - p(n))^2 + p(n)(1 - p(n)) \right) + \left(\frac{\lfloor n^2 t \rfloor - \lfloor n^2 r \rfloor}{n} p(n) \right)^2.
\end{aligned}$$

If $\lfloor n^2 t \rfloor - \lfloor n^2 s \rfloor \geq 2$, $\frac{\lfloor n^2 t \rfloor - \lfloor n^2 r \rfloor}{n^2} \leq 2(t - s)$ and $\frac{\lfloor n^2 r \rfloor - \lfloor n^2 s \rfloor}{n^2} \leq 2(t - s)$. If $\lfloor n^2 t \rfloor - \lfloor n^2 s \rfloor \leq 1$, then $\lfloor n^2 t \rfloor - \lfloor n^2 r \rfloor = 0$ or $\lfloor n^2 r \rfloor - \lfloor n^2 s \rfloor = 0$. Therefore, we can find out a constant $C > 0$ such that

$$\Gamma(r, s, t) \leq C[t - s]^2.$$

By application of Theorem 13.5 in [4], we conclude the convergence (5). \square

5. CONCLUSION AND COMMENTS

We might also think of another related model, that is, a spatial population model in which children either occupy the same location as their parents or migrate to new places and start growing colonies on their own. If the distances of migrations and the sizes of immigrants are random, it seems that we need to consider a process indexed by the universal tree such that there is a random variable for each edge and vertex. We might associate it with a CSBP in continuous time.

Acknowledgements

I would like to thank M. Yor for his advice and help, including introducing Bertoin's paper to me and showing some useful arguments to me.

REFERENCE

- [1] K.B. Athreya and P.E. Ney. *Branching processes*. Springer-Verlag, Berlin, 1972.
- [2] J. Bertoin. The structure of the allelic partition of the total population for Galton-Watson processes with neutral mutations. *Ann. Probab.*, 37:1502–1523, 2009.
- [3] J. Bertoin. A limit theorem for tree of alleles in branching processes with rare neutral mutations. *Stochastic Processes and their Applications*, 120:678–697, 2010.
- [4] P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics. Wiley-interscience, second edition edition, 1999.
- [5] M. Jiřina. Stochastic branching processes with continuous state space. *Czechoslovak Math.*, 8:292–313, 1958.
- [6] O. Kallenberg. *Foundations of modern probability*. Probability and its Application. Springer-Verlag, New York, second edition, 2002.
- [7] J. Pitman. Combinatorial stochastic processes. *Lect. Notes in Maths.*, 1875, 2006.
- [8] H.S. Wilf. *Generatingfunctionology*. Academic Press, 1994.